



Questions and Answers

Author(s): Ernest Rubin

Source: *The American Statistician*, Vol. 8, No. 5 (Dec., 1954), pp. 19-21

Published by: American Statistical Association

Stable URL: <http://www.jstor.org/stable/2681548>

Accessed: 06/01/2010 14:35

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=astata>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician*.

<http://www.jstor.org>

QUESTIONS AND ANSWERS

Edited By ERNEST RUBIN

*U. S. Department of Commerce
and American University*

Statistical Relationships and Proof in Medicine

Within the last two years a formidable controversy has developed as a result of investigations concerning smoking and lung cancer. Statistical data indicate that smokers have a higher incidence of lung cancer than non-smokers. The experience with the smoking—lung cancer controversy suggested the following question: Are there instances in the history of medicine in which a statistical association proved causation or has the proof of a causal relationship in medicine always depended on direct experimentation? I wish to thank Mr. Jerome Cornfield of the Office of Biometry, National Institutes of Health for preparing the following analysis of this question.

There are several preliminary issues raised by the question that need prior discussion. The first, the concept of proof, need not detain us long. Proof has a well-defined meaning in mathematics, but not in empirical science. The truth of a mathematical proposition can be demonstrated; the evidence for an empirical proposition, i.e., a statement in natural science, can be made strong or even overwhelming (despite the apparent impossibility of a satisfactory calculus of evidence). It is doubtful, however, if such propositions can ever be regarded as proved. New evidence (e.g., the discovery of black swans) may cast an entirely different light on a well-established proposition, and in an empirical science, as opposed to mathematics, there are no postulate systems which delimit the kind of new evidence that can be found. If we ask for proof in medicine, or any other empirical science, we may be asking for something that does not exist.

A second issue raised by the question is the exact nature of the distinction between a relationship based upon a statistical association and one based on direct experimentation. We all have a vague feeling that if we can make an event occur, we understand it better than if we simply observe it passively. On analysis this feeling seems to reduce to two propositions like the following: We are initially skeptical of any relationship based upon simple observation because the effects of other possibly important variables are not controlled and may account for the observed association. We are initially impressed by any relationship established by experiment because we feel that the effects of other important variables are controlled and cannot account

for the association. The distinction we feel between a relationship based upon a statistical association and one based upon direct experimentation is thus a distinction between relationships that may be explained by other variables and those that cannot.

Although this statement may formalize our intuitions it is an oversimplification of the actual facts. First, there are cases in which uncontrolled observations can be so analyzed as to eliminate the possibility that extraneous variables account for the observed association. The classical example of this is Snow's demonstration (1) in the middle of the nineteenth century, before the birth of bacteriology, that cholera was transmitted through polluted water. Even the most skeptical critic cannot quarrel with the conclusions drawn from his observations on the clustering of deaths about a particular source of polluted water, the famous Broad Street pump; particularly after his demonstration that mortality from cholera among subscribers of a water company that drew its supply from the Thames River was 14 times as high as that among subscribers of the competing company whose water was sewage-free. The official inquiry which followed agreed that "fecalized drinking-water . . . may breed and convey the poison [of cholera]" although with a caution that is perhaps not peculiar to the Victorian era added, "[so would] fecalized air."¹ Nor do we have to go back 100 years to find examples in which the effects of specific extraneous variables were eliminated from observational material by methods short of direct experimentation. Cross-classification of observations is an obvious, but often surprisingly powerful method of accomplishing this, for some recent examples of which references (3, 4, 5 and 6) may be instructive.

Secondly, our intuitions may be misleading because there is no automatic guarantee in any particular instance that extraneous variables have been controlled by direct experimentation. This may seem to deny the

¹ It is not entirely irrelevant to recall at this point the experience of Max von Pettenkofer who, many years later, to prove beyond any doubt that water-borne bacteria did *not* cause cholera, drank, and induced several of his students to drink, a whole glass full of the bacilli. They not only all survived but reported nothing worse than a bellyache (2).

great virtue claimed for randomization, the automatic balancing out among treatment groups of the effects of other variables, whether or not we are aware of their existence. The denial is more apparent than real, however. Consider for example an experiment designed to study the effect of removing an organ, say the thyroid gland, on some biological response, say blood sugar level. We may randomize animals among a control and thyroidectomized group and thus eliminate in the usual probability sense the possibility that any large difference between the two treatments arose from the different characteristics of the animals treated. But we have not eliminated the possible effect of other extraneous variables in which the experimenter is equally interested such as the operation removing the thyroid, or the non-specific effect of thyroidectomy on weight loss. While it is perhaps possible to control these specific variables, for example, sham operation and under feeding the sham operated controls might be regarded as providing such a control (7), randomization by itself is insufficient. We must indicate the specific variables we wish to control and must devise specific experimental procedures to control them.

Having thus argued that there is no difference in kind between the two types of evidence it is of course necessary to add that there is a very important difference in degree. It is a good deal more difficult to control variables in observational than in experimental material, so that the experimental method has unravelled and will continue to unravel mysteries before which uncontrolled observation would be powerless. But there is no difference in principle. There are no such categories as first-class evidence and second-class evidence. There are merely associations, whether observational or experimental that, in a given state of knowledge, can be accounted for in only one way or in several different ways. If the latter, it is our obligation to state what the alternative explanations or variables might be and to see how their effects can be eliminated, while if the former it is equally our obligation to state so. To distinguish between statistical association on the one hand and relationships that are established by experimentation on the other, without any reference to alternative variables that are present in one case but not the other, seems to us to be neither good statistics, good science, nor good philosophy—though it may be good red herring.

If we consider the tobacco-lung cancer question, for example, one possible set of extraneous variables that might explain the higher incidence for smokers are those arising from self-selection. Thus, some small proportion of the population, say 5 percent, might have some special trait (or traits), say high blood levels of certain hormones, which both initiate lung cancer and make the possessors smoke. Of the remaining adult male population, 75 percent smoke for other reasons

and do not develop lung cancer. It is possible to conceive but impossible to conduct an experiment that could settle this question. A large group of adolescents would be allocated at random to different smoking groups, compelled to remain on the assigned smoking schedule, followed for the 30 to 60 years required for lung cancer to develop and the lung cancer incidence computed for each group. This and, as nearly as one can see, only this, would entirely eliminate self-selection as an explanation. Short of this one must rely upon indirect evidence. If self-selection were the complete explanation of the difference, then tobacco smoke would not be a carcinogen for human lung tissue. One might consequently investigate this question by asking is it a carcinogen for any other type of tissue that one can reasonably experiment with, say human or mouse skin? If the answer had been no, this might have been regarded as some type of evidence for the self-selection hypothesis, although no one would regard the evidence as very strong. The recent induction of skin tumors in mice by tobacco tars (8) might similarly be considered evidence against the self-selection hypothesis, but again far from strong. In any event the recent announcement by the Tobacco Industry Research Committee that it would investigate psychological differences between smokers and nonsmokers suggests that we have not heard the last of the self-selection hypothesis. No matter what one's opinion on the plausibility of this as an explanation,² the actual investigation of whether specific differences that might arise from self-selection do in fact account for the association could be constructive, even if the results obtained were negative.

This discussion of preliminary issues³ in one sense also disposes of the main question, but the history of medicine on this point is interesting and a few words may be in order. There are numerous instances that one can cite in which the most important source of medical knowledge on a subject was supplied by statistical associations. Thus, the observation that there is a close inverse association between the amount of natural fluorides in water and the amount of dental caries among children drinking it, has induced numer-

² It is easy to sympathize with, even if one cannot entirely share the exasperation expressed by Greenwood and Yule on a related point (9). "[The vaccinated group] may all have been vegetarians, or nonsmokers, or red-headed, and all or any of these things may render them less likely to contract cholera; but we do not see why objections which no sensible man would allow to influence him in the ordinary affairs of life should suddenly acquire scientific importance when the question is one of interpreting statistics."

³ There is one additional preliminary issue that deserves mention. The phrasing of the question suggests that its framer subscribes to the somewhat old-fashioned view that it is either possible or desirable to discuss knowledge without any reference to the possible actions to which it will lead. I have not challenged this view only because I share it.

ous municipalities to add fluorine to their water supply. The resulting decline in the incidence of dental caries in these municipalities may be considered a "direct experimental proof" of the proposition that fluorine inhibits the development of caries, but the fluorine was not added in order to study this question experimentally but rather to bring about a result indicated by the associations. It is true that the results of adding fluorine to the drinking water of experimental animals also pointed in the same direction (10), but as evidence this apparently was not given much weight. Shaw, for example, in his excellent summary of the subject (11) does not even mention the results with experimental animals.

In the study of the effects of therapy the application of modern ideas of experimental design is a very recent development, for an account of which the reader is referred to Hill's very interesting article (12). In recent years there have been several well-conceived experiments to test the efficacy of different preparations, such as gamma globulin, in protecting against the subsequent development of disease. But methods that were established in the past such as vaccination against smallpox, have never received such a carefully controlled experimental test. There are of course dozens of studies to show that individuals who had been voluntarily vaccinated developed less smallpox than others, and that when they did develop it, the outcome was less frequently fatal. But none of these studies ruled out the possibilities of self-selection any more effectively than they are now ruled out in tobacco-lung cancer studies.

In the study of infectious disease there are naturally almost no examples of direct experimental demonstration on humans. Walter Reed's experiments on yellow fever are well known, but it is difficult to find other cases. Perhaps the nearest is the ghastly episode that occurred in Lübeck in 1926, when out of 249 babies accidentally inoculated with enormous numbers of living virulent tubercle bacilli, 76 died (13). If one is willing to overlook the absence of a placebo-inoculated control group, and refrains from asking, "if the bacilli cause tuberculosis, why didn't all the inoculated children develop the disease?", this perhaps is "proof of a causal relationship." (The 173 Lübeck babies who did not die developed only minor lesions and were still free of tuberculosis when last observed 12 years later.)

In short, if we insist on direct experimental demonstration on humans there are many widely held beliefs that must be regarded as without solid foundation. If we believe that vaccination protects against the development of smallpox it is not because there has been a direct experimental demonstration but rather (a) there is a good deal of evidence that is consistent with this hypothesis, and (b) over the course of many

years no evidence has been produced to support any alternative hypothesis. The truth of the matter appears to be that medical knowledge (and, one suspects, many other kinds as well) has always advanced by a combination of many different kinds of observation, some controlled, and some uncontrolled, some directly and some only tangentially relevant to the problems at hand. Although some methods of observation and analysis are clearly to be preferred to others when a choice is possible, there are no magical methods that invariably lead to the right answer. If we cannot specify exactly what has been learned in medicine from the study of statistical associations, we can at least say that we could not have accumulated the knowledge we have without them.⁴

REFERENCES

1. J. Snow, Snow on Cholera, being a reprint of two papers by John Snow, M.D., 1936, The Commonwealth Fund, New York.
2. R. H. Shryock, The Development of Modern Medicine, 1947, Alfred A. Knopf, New York, p. 292.
3. W. E. Heston, M. K. Deringer, I. R. Hughes and J. Cornfield, Interrelation of specific genes, body weight and development of tumors in mice, 1952, *J. National Cancer Institute*, 12: 1141.
4. E. L. Wynder, J. Cornfield, P. D. Schroff, K. R. Doraiswami, A Study of environmental factors in carcinoma of the cervix, 1954, *Amer. J. Obst. and Gyn.*, 68: 1016.
5. W. W. Smith, R. Q. Marston, H. J. Ruth and J. Cornfield, Granulocyte count, resistance to experimental infection and spleen homogenate treatment in irradiated mice, 1954, *Amer. J. Physiol.*, 178: 288.
6. W. W. Smith, L. Gonschery, I. M. Alderman and J. Cornfield, Effect of granulocyte count and litter on survival of irradiated mice, 1954, *Amer. J. Physiol.*, 178: 474.
7. R. O. Scow and J. Cornfield, Effect of thyroidectomy and food intake on oral and intravenous glucose tolerance in rats, 1954, *Amer. J. Physiol.*, 179: 39.
8. E. L. Wynder, E. A. Graham, A. B. Croninger, Experimental production of carcinoma with cigarette tar, 1953, *Cancer Res.*, 13: 855.
9. M. Greenwood, Epidemics and Crowd Diseases, 1937, Macmillan Co., New York, p. 96.
10. I. Zipkin and F. J. McClure, Inhibitory effect of fluoride on tooth decalcification by citrate and lactate in vivo, 1949, *J. Dental Research*, 28: 151.
11. J. H. Shaw, Should fluorides be added to public water supplies?, 1954, *Scient. Monthly*, 79: 232.
12. A. B. Hill, The Clinical trial, 1952, *New England J. of Med.*, 247: 113.
13. R. and J. Dubos, The White Plague, 1952, Little, Brown & Co., Boston, p. 122.
14. A. B. Hill, Observation and experiment, 1953, *New England J. Med.* 248: 995.

⁴ After completing this answer my attention was called to Hill's Cutler Lecture on Preventive Medicine (14) in which much the same issues that we have covered were also considered—in, however, a more comprehensive, lucid (and reasonable) manner. The reader is enthusiastically referred to it if he is at all interested in pursuing the subject.